7. NUMIDENT FILE 7.2

7.1. THE NUMIDENT MASTER FILE 7.2
7.1.1. Source of Data and File Structure 7.2
7.1.2. Crosswalk of Research Subject and Data Elements 7.4
7.1.3. Data Usage Issues 7.5
7.2. DESCRIPTION OF NUMIDENT DATA ELEMENTS 7.5
7.2.1. CITIZEN Citizenship Code 7.6
7.2.2. CYCDTE Enumeration Cycle Date 7.7
7.2.3. DOB Date of Birth 7.8
7.2.4. DOD Date of Death 7.9
7.2.5. ENTCD Entry Code 7.10
7.2.6. POB City and/or County of Birth 7.12
7.2.7. POBFORID Foreign Place of Birth Indicator 7.13
7.2.8. POB-STATE-COUNTRY State or Country of Birth 7.14
7.2.9. RACE Ethnic Description 7.15
7.2.10. SEX Sex Code 7.16
7.2.11. SSN Social Security Number 7.17
7.3. NUMIDENT-BASED EXTRACTS 7.19

7. Numident File

7.1. The Numident Master File

Until the early 1970s, data on SSA clients were kept only in hardcopy form. In the early 1970s, SSA began to create a central database for client information called the Numident. The Numident is a numerically-ordered master file of assigned Social Security Numbers (SSNs).). The official repository of death information for the Social Security Administration is the Death Master File (DMF), stored on Numident and based on the date received from the DACUS operation. The Numident file receives information from SSA and other sources including all State and Territorial Bureaus of Vital Statistics and the Veteran's Administration. The Numident became fully electronic in the 1980s.

Researchers tend to use the Numident to obtain demographic data for survey participants. The Numident is also used for mortality studies, since it carries birth and death dates. However, The Master Beneficiary Record (MBR) is considered to be the most accurate record for mortality, if one exists for the number holder.

Alpha-Index Master File (Alphident)

The Alpha Index Master File is an alternative version of the Numident. This file was created to search for an SSN when only minimum client-information is available and the SSN is unknown. The extract includes client first, middle, and last name, DOB, sex, mother's first and last name, father's first and last name, and SSN. Field office and central office staff can access this file.

Because the Alpha Index Master File is a subset of the Numident, this chapter focuses only on the Numident.

7.1.1. Source of Data and File Structure

There is one Numident record for each SSN ever issued. The record houses the identifying information given by the applicant for an SSN ordered by the date of the application. Each Numident record can have up to 300 sections or "entries." Each entry represents an addition or change to the information pertaining to that SSN. For example, a name change results in reissuing a new Social Security card with the new name and the old SSN. A new entry, rather than a new record, is added to record that information.

Entries are of fixed length, and the length of Numident records depends on the number of entries. In 1999, there were approximately 400 million records and approximately 725 million entries on the Numident.

There are five different types of entries, corresponding to their informational source. The names of the entry types as defined in the Numident data dictionary are NSS5, NCLM, NCOR, NDTH, and NESI. These names are mnemonic words rather than abbreviations and refer to SS-5 forms (NSS5), claims (NCLM), correspondence (NCOR), death (NDTH), and enumeration special indicators (NESI). With the exception of NDTH entries, entries are not overwritten and individual entry fields are never deleted, except in rare circumstances. Instead, new entries are created with updated information. This implies that multiple entries may need to be searched for the most recent information.

NSS5: Client Identification Information

NSS5 entries include, for example, initial SSN assignments, name changes, and DOB changes. The majority of entries on the Numident are NSS5 entries. NSS5 entries are built from data collected in field offices and from data provided from State vital statistics through enumeration at birth. The Modernized Enumeration System (MES) is one of the SSA systems used in field offices to enter NSS5 data. The MES writes to the Integrated Client Database (ICDB). ICDB is an attempt to create a single view of critical elements that identify clients. Critical elements include names, date and location of birth, gender, date of death, and parent names. This allows standardization of client-identifying information across systems that are used to pay benefits for SSA-administered programs.

NCLM: Information on Title II Claims

NCLM entries were built when SSA moved from paper to computerized systems. Prior to the Numident, only the most recent paperwork was maintained in paper files for each record. Consequently, when the Numident was created, the only data available for many records was from claims paperwork. NCLM entries are no longer created. NCLM entries on the Numident contain minimal data and carry information on individuals who are mostly deceased. It is sometimes considered a “catch-all” entry; for example, in a small percentage of records the date of death can only be found in the NCLM entry. In fewer than 1 percent of the records, NCLM data may not be reliable because of how older claims were processed. For example, a widow’s identification may have been put in the wage-earner’s identification fields because the widow may never have applied for her own SSN.

NCOR: Client Correspondence Information

NCOR entries are built when a client writes to SSA about any problem.

NDTH: Information on Client Death, Including Date of Death

Only one NDTH entry is allowed for each SSN. Unlike with other entries, updated information for these entries may be rewritten or deleted. Processing of death data began with DACUS in 1988; death information for a large percentage of people who died prior to this date may not be posted. Data for the NDTH entry come from many sources, including the Veteran's Administration, State and Territorial Bureaus of Vital Statistics, SSA field offices, other SSA administrative files, etc. The vast majority of entries come from vital statistics; this reporting began in February 1988. NDTH entries are built through the DACUS system. Entries with death information are created from incoming sources and sent to DACUS. DACUS runs every night after all other computer processes. DACUS compares alleged date of death against the MBR to determine whether the client was paid after the alleged date of death. If there are no discrepancies with the MBR data the information goes to the Numident for posting. Since only one death entry is allowed, a complex prioritization scheme exists to compare information coming through DACUS with any existing NDTH entry on the Numident. If discrepancies are found between incoming data and existing data, an investigation will be conducted to reconcile the discrepancies. If incoming data are considered better, the system will overwrite the existing data with the new data. If an NDTH entry is built for a client who is later determined to be living, the NDTH entry is removed.

NESI: Special indicator entries

NESI entries include information for records that are flagged for an investigation of some sort, such as potential fraud cases. These records are added to existing Numident records to indicate, for example, fraudulent identity or illegal alien status.

7.1.2. Crosswalk of Research Subject and Data Elements

The Numident by itself is generally used to verify SSN, name, and date of birth or death for demographic studies involving births and mortality, if an MBR record is not available. For analyses based on other SSA master files, the Numident usually provides the most recent reliable source for individual birth dates, death dates, gender, location of birth, and ethnicity, although ethnicity is supplied by the applicant and is not a mandatory item to be filled out. Thus, researchers should request that items such as birth date, gender, and death date be pulled from the MBR first, and data pulled from the Numident for those individuals without a MBR record. Many of the extracts that are produced already pull the MBR first and use the Numident as a secondary source of data verification.

7.1.3. Data Usage Issues

Multiple SSNs

SSA currently uses matching procedures to help ensure that clients are not assigned multiple SSNs. However, in past years, some clients were assigned multiple SSNs. For instance, clients would apply for a new SSN with each job change. SSA maintains a file called the MULTX, which contains a table of multiple numbers allegedly assigned to the same client. This file is independent of the Numident.

Data Output from Numident

Output data may come from one or more entries or entry types on the Numident. When SSNs are run against the Numident, the finders program (which looks up specified SSNs) delivers the complete data for the SSN, i.e., every entry. When SSA Numident analysts are looking for data, they generally follow a hierarchy based on the type of information requested and entry type. In checking for all but death data on the Numident, the NSS5 entry is always used first. Analysts will select data from the most recent cycle date possible. If the information does not exist in NSS5 entries, the analysts will look at the NCLM entry. For death data, the NDTH entry is always checked first, followed by the NCLM entry.

7.2. Description of Numident Data Elements

The following pages contain the list of data elements, or fields, in the Numident master file that have been viewed as the most useful to the research community. Several elements will occur in multiple types of entries. Elements are always located in the same position across the different entry types.

The table below lists the data elements covered in the data dictionary. They are presented in alphabetical order. The page headers note the data element that is being described on a given page.

7.2.1. CITIZEN Citizenship Code

DESCRIPTION AND PURPOSE

This data element indicates the SSN holder's U.S. citizenship or alien status.

Since May 1981, it identifies the number holder's status as a U.S. citizen or work status for non-citizens.

POSSIBLE VALUES

Character 1 position
A – U.S. citizen
B – Legal alien, authorized to work
C – Legal alien, not authorized to work
D – Other
E – Alien student—restricted work authorized
F – Conditionally legalized alien

USAGE NOTES

Effective 8/97, the Modernized Enumeration System (MES) no longer permits the codes of 'E' or 'F' to be added to Numident. 'E' and 'F' categories now fall under 'D.'

Category C supports administrative uses, such as drivers' licenses, which require an SSN for identification purposes. Some aliens are not able to work but may still hold a permit to drive. Otherwise, this field is used to determine work eligibility.

7.2.2. CYCDTE Enumeration Cycle Date

DESCRIPTION AND PURPOSE

This data element indicates the date an enumeration entry was added to the Numident.

It identifies when the data for the particular entry was posted to the Numident. For records prior to the mid-1970s, this represents the date written on the SS-5 form.

POSSIBLE VALUES

Character 8 position
Blank - Default
XXXXXXXX - Complete date is unknown
CCYYMMXX - Acceptable for unknown day (DD)
CCYYMMDD - Correct entry

USAGE NOTES

This data element is stored in CCYYMMDD format for entry-sorting purposes.

It is displayed in MMDDCCYY format on queries

Only dates from NDTH are cleared or reset.

Entries are sorted within record by CYCDTE.

This data element is used to determine the most current data.

7.2.3. DOB Date of Birth

DESCRIPTION AND PURPOSE

This data element indicates the month, day, century, and year an individual was born.

It is used to help identify an individual and may be used in searching for his/her SSN on the related Alpha-Index (Alphident) file.

POSSIBLE VALUES

Character 8 position
Blank - Default Value
XXXXXXXX - Complete date is unknown
MMXXCCYY - Acceptable for unknown day (DD)
CCYYMMDD - Correct entry

USAGE NOTES

Only dates from NDTH are cleared or reset.

7.2.4. DOD Date of Death

DESCRIPTION AND PURPOSE

This data element indicates the client's date of death as reported to SSA.

The death entry is the official SSA record of death; therefore, this is the client's official date of death.

POSSIBLE VALUES

Character 8 position
Blank - Default Value
XXXXXXXX - Complete date is unknown
MMXXCCYY - Acceptable for unknown day (DD)
MMDDCCYY – Correct format

USAGE NOTES

Only dates from NDTH are cleared or reset.

Only one date of death allowed on file; previous dates are overwritten.

As per agreement, states need only provide month and year of death. However, some states also provide day.

Both the NDTH and NCLM entries have a DOD field. Therefore, theoretically, DOD may occur on the Numident record more than once. However, in reality, DOD will most likely exist on only one of these entries (usually the NDTH) and will not be repeated within the entry. Changes to the NDTH entry will overwrite the old NDTH data.

If trying to determine the official Date of Death, search engines first look at the NDTH entry. If that entry has no data, then the NCLM entry is reviewed.

Most information will come from NDTH entry.

There are 3.3 million Numident records with DOD on both the NCLM and NDTH entries.

7.2.5. ENTCD Entry Code

DESCRIPTION AND PURPOSE

This data element identifies the type of entry, e.g., NSS5 and NDTH. Further, the specific code tells the source of the data.

POSSIBLE VALUES

Character 1 position

NSS5 Entry:

G – SS5 generated by using the MEF cross-reference (MULTX), or MBR files (no longer used)
J – New account established prior to central office issuance of SSNs
K – Duplicate name change
P – Change in individual's identifying information (replaced by "2" below)
S – Supporting application for SSN card
V – Voided SSN
W – Welfare request (no longer used)
0 – Original application for SSN card
1 – Date of birth correction
2 – Name change, duplicate SSN, replacement card issued, or correction of misspelled name
4 – Reestablish voided SSN
5 – Sex correction
9 – Internal correction

NCLM Entry:

D – Death claim (claim for benefits made by a survivor)
L – Life Claim (claim for benefits made by a living client)

NCOR Entry:

Q – Correspondence

NESI Entry:

E – ESI entry

NDTH Entry:

T – Death entry

USAGE NOTES

Only codes from NDTH are cleared or reset.

As of 1999, approximately 47 percent of all entries were '0', approximately 35 percent were '2'

'G' and 'W' are no longer added, and 'P' is replaced by '2'.

'G' – part of 1986 study where SS5 entries were generated based on data search of other Administrative files.

'J' – values assigned in "emergencies" (e.g., welfare cases needing immediate SSN).

'S' – not all information could fit on SS5 form; additional paperwork is stapled to SS5.

7.2.6. POB City and/or County of Birth

DESCRIPTION AND PURPOSE

This data element indicates the name of the city and/or county where an individual was born.

It is used to help identify an individual.

POSSIBLE VALUES

Character 13 position

Actual county/city name.

Valid data appears in positions 1-12; position 13 is set to set to '*' as an overflow indicator.

USAGE NOTES

In general county name is seen when an individual is not born in a city. In some instances, this data element has both city and county.

7.2.7. POBFORID Foreign Place of Birth Indicator

DESCRIPTION AND PURPOSE

This data element provides a simple indicator used to identify persons born outside the United States.

POSSIBLE VALUES

Character 1 position

* - Born outside the United States
Blank – Default value

USAGE NOTES

This data element must be used with the POB-State-Country data element to determine state and country of birth.

7.2.8. POB-STATE-COUNTRY State or Country of Birth

DESCRIPTION AND PURPOSE

This data element indicates the U.S. state, U.S. territory, or foreign country where an individual was born.

It is used to help identify an individual and may be used in search for his/her SSN on the related Alpha Index file.

POSSIBLE VALUES

Character 2 position

Blank–Default Value

Valid abbreviations for states and countries can be found in POMS RM 0029.220

USAGE NOTES

Some codes have been changed (e.g., PR has been replaced by RQ, all territories now have the letter Q in the second character, and pre- and post-war Germany have two separate codes).

Codes may not be unique. For example, code CA represents both California and Canada. One therefore needs to use the Foreign Country Indicator to confirm country.

Some foreign country codes for records established prior to 1997 may be incorrect because edits were not applied prior to 8/97.

7.2.9. RACE Ethnic Description

DESCRIPTION AND PURPOSE

This data element indicates ethnicity, as voluntarily provided by the applicant.

Officially, this information is collected for statistical purposes only and for studies of demographics by agencies such as the Bureau of the Census.

POSSIBLE VALUES

Character 1 position

0 – Unknown
1 – White
2 – African American or African origin
3 – Other
4 – Asian or Pacific Islander
5 – Hispanic (obsolete)
6 – North American Indian or Eskimo (obsolete)

USAGE NOTES

This field is optional on input by the applicant.

The input screen in the field offices where this information is generally entered lists different values.

Prior to 1982, ethnicity data were not collected. From 1982 through 1988, collection was mandatory. In 1989, it became optional.

7.2.10. SEX Sex Code

DESCRIPTION AND PURPOSE

This data element indicates the gender of the number holder. It is used to help identify an individual and may be used in searching for his/her SSN on the related Alpha Index File.

POSSIBLE VALUES

Character 1 position

On the Numident:

0 – Unknown
1 – Male
2 – Female

On the Alpha Index:

U – Unknown
M – Male
F – Female

USAGE NOTES

Only codes from NDTH are cleared or reset.

Although this data element has always existed on the file, it has not always been mandatory to collect this information; thus, this data element may be listed as unknown.

Unknowns also arise because, at time of data transfer from hardcopy, information may have been illegible.

7.2.11. SSN Social Security Number

DESCRIPTION AND PURPOSE

This data element provides a number assigned by SSA to uniquely identify a client.

Within SSA, the SSN is to control:

- The record of all earnings the client has in FICA-covered or MQGE-covered employment.
- Claims for social security benefits.
- Claims for supplemental security benefits.
- Claims for Medicare benefits.

POSSIBLE VALUES

Character 9 position

XXX-YY-ZZZZ

Area: XXX (001-999). Assigned based on geographic area of the U.S.
Exceptions include:
700s - Railroad Retirement beneficiaries. Some of the 700s are now being released to states that have used up their initial allocation.
800s – Not yet used for SSNs.
900s – Not yet used for SSNs. Some numbers in this series were "pseudo numbers" given to welfare benefits (1974-early 1990s). These are not included in the Numident file but exist on other administrative files.
Group: YY (01-99). Time of issue: Even numbers are assigned first (02-98), followed by odd numbers (01-99).

Serial: ZZZZ (0001-9999). Serial number.

USAGE NOTES

Only SSNs from NDTH are cleared or reset.

Non-program users of the SSN for record control include:

State Unemployment Insurance Agencies;
The Office of Personnel Management;
The Internal Revenue Service;
Schools;
Public Assistance Agencies;
The Department of Defense;
The Department of Veteran's Affairs;
The Indian Health Service;
The Parent Locator Service;
Motor Vehicle Bureau.

SSN were not created to be a universal identifier, but they are increasingly being used by the private sector as an identification tool.

7.3. Numident-Based Extracts

The Numident has no associated recurring extracts. The Alpha Index Master File (sometimes called Alphident) contains a subset of Numident data. The access key is a combination of SOUNDEX key, client's name, and date of birth. It is not accessible to the research community.